

Context-aware pattern discovery for moving object trajectories

Mohammad Sharif, Ali Asghar Alesheikh, and Neda Kaffash Charandabi

Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran; msharif@mail.kntu.ac.ir, alesheikh@kntu.ac.ir, n_kaffash@yahoo.com

Abstract: Movement of point objects are highly sensitive to the underlying situations and conditions during the movement, which are known as contexts. Analyzing movement patterns, while accounting the contextual information, helps to better understand how point objects behave in various contexts and how contexts affect their trajectories. One potential solution for discovering moving objects patterns is analyzing the similarities of their trajectories. This article, therefore, contextualizes the similarity measure of trajectories by not only their spatial footprints but also a notion of internal and external contexts. The dynamic time warping (DTW) method is employed to assess the multi-dimensional similarities of trajectories. Then, the results of similarity searches are utilized in discovering the relative movement patterns of the moving point objects. Several experiments are conducted on real datasets that were obtained from commercial airplanes and the weather information during the flights. The results yielded the robustness of DTW method in quantifying the commonalities of trajectories and discovering movement patterns with 80% accuracy. Moreover, the results revealed the importance of exploiting contextual information because it can enhance and restrict movements.

Keywords: Pattern discovery, Similarity measurement, Trajectory, Context

1. Introduction

Progress in sensing, navigation, and tracking technologies, such as global navigation satellite system (GNSS), multi-functional sensors, and so forth, enable access to unprecedented amounts of moving point object (MPO) data. The sequence of a tracked point object represents the spatial and temporal history of it, which is known as trajectory (Spaccapietra, Parent, & Spinsanti, 2013). However, the goal is to not only collect spatio-temporal information of objects' movements but also to generate knowledge from their trajectories. Therefore, analyzing the spatial behaviors of MPOs by studying their trajectories and discovering their movement patterns have been of great interest to the cartography and GIScience community recently.

Normally, movement pattern discovery can be fulfilled through similarity measurement of trajectories. However, in spite of the recent progress in the similarity measure of trajectories, most of the existing methods deal only with the positional information of the MPOs over time, and very little attention has been paid to the influential factors, which are known as *contexts*, during the move. Movements are highly affected by internal and external contexts in real world applications. The former is any factor that is related to the MPO's characteristics, states, and conditions (e.g., intension, mode), while the latter is dedicated to the geographical and environmental conditions (e.g., weather condition, landscape) during the move (Nathan et al., 2008). Both types of contexts cause MPOs to react differently according to their variations.

On the one hand, a review of the relevant literature reveals a few research contributions that have solely exploited the external (geographic) context information in the trajectory similarity measure process. For example, a conceptual model for context-aware movement analysis was proposed based on events and was validated by exploring the relation of moving objects (roe deer and lynxes) with each other and moving objects to their geographic context (open areas) (Andrienko, Andrienko, & Heurich, 2011). Buchin and her colleagues incorporated geographic context fac-

tors in a trajectory similarity measurement process (Buchin, Dodge, & Speckmann, 2014). Their proposed method was based on Hausdorff, Fréchet, and equal time distances and ran by the subdivision modeling of context. In that research, contexts were modeled as a polygonal subdivision of land use (sea/land) and meteorological parameters for a North Atlantic hurricane dataset and albatross tracking data in Euclidean space.

On the other hand, discovering patterns among trajectories based on their similarity search results has received less attention, so far. According to related research on pattern mining of MPOs, Dodge and her colleagues used the Euclidean distance approach for a similarity measure of trajectories while aiming to find the so-called coincidence (spatio-temporal) and concurrence (movement parameters) patterns of North Atlantic hurricanes (Dodge, Weibel, & Laube, 2011). From the methodological standpoint, the employed Euclidean distance is a very brittle distance measure because it can be applied only to trajectories of the same size (length). Therefore, a robust alternative method that is capable of accommodating trajectories that have different sizes is crucial. From dimensional perspective, this study covers only the space, time, and a number of movement parameters in pattern mining among trajectories.

To the best of our knowledge, no other research has been conducted in incorporating both internal and external contextual information for similarity measure and pattern discovery of trajectories. This research contributes to fill this gap by contextualizing the pattern discovery of multi-size trajectories while incorporating both internal and external contextual information along with positional data in similarity measurement of trajectories.

2. Methodology

This section first focuses on the state-of-the-art conceptualization of context as the basis for contextualizing trajectory analysis. Second, it describes the method by which the similarities of trajectories are going to be contextually quantified. Third, the pattern discovery procedure is explained. The key elements of the methodology are illustrated in Figure 1 and more descriptions are provided in the rest of this section.

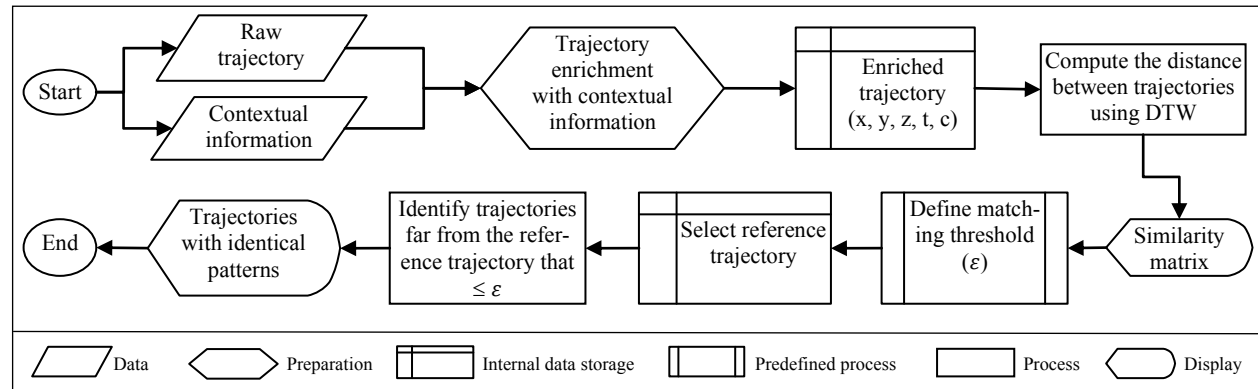


Fig. 1. Methodology for contextualizing similarity measure and pattern discovery for trajectories.

2.1 Context

The term *context* has a wide range of meanings with regard to the variety of applications and research domains. Various definitions, descriptions, and classifications have been provided for context in the literature. In this research we adhere to Dey's comprehensive definition of context as "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves*" (Dey, 2001). Context is about the entire situation (current state) relevant to movement. It specifies the elements that must be collected to model a situation (Crowley, Coutaz, Rey, & Reignier, 2002). Context provides information regarding the present status of people (in-

dividuals, groups), places (indoors, outdoors), and things (physical objects, computing devices) in the environment. For example, an individual's situation can be characterized by the contexts of the individual's profile, the individual's surroundings, the individual's social interactions, the individual's activity, etc. Consequently, context is categorized into (1) internal context, which consists of all of the properties that directly relate to the MPOs, and (2) external context, which refers to any outer factor that influences the process of movement extrinsically.

2.2 Contextualizing the similarity measure of trajectories

The majority of research contributions have focused on spatial or spatio-temporal similarity measures of trajectories in 2- or 3-dimensional spaces (x, y, t). In this research we add another dimension to trajectories, called *context dimension* (c), in the sense that each element of trajectory can be annotated by D dimension. The D dimension, with regard to the definition of context, can be labeled as any quantitative value (e.g., speed, direction) of internal and external contexts, and be represented as (x, y, (z), t, c).

Several similarity measure methods have been introduced and developed to gauge the commonalities of MPOs' trajectories. We focus on a method that belongs to the time series family, which is known as dynamic time warping (DTW). Its initial version was proposed for speech recognition by Sakoe and Chiba (Sakoe & Chiba, 1978), and has been commonly employed in diverse fields of science. The DTW method relies on pair-wise comparisons of trajectory sampling points and warps the time to measure the similarities of trajectories. DTW is suitable for matching trajectories that have missing information, supports local time shifting, addresses parametric data very well, and handles dissimilar trajectory lengths (sizes). In contrast to Euclidean distance, DTW enables trajectories to be compressed or stretched along the temporal dimension to measure their match non-linearly. The computational time of DTW is $O(n * m)$, which directly relates to the lengths of the trajectories.

For two trajectories $S[s_1, s_2, \dots, s_n]$ and $T[t_1, t_2, \dots, t_m]$ of lengths n and m , respectively, the DTW distance is calculated as follows:

$$\begin{aligned} DTW(\varphi, \varphi) &= 0 \\ DTW(s_i, \varphi) &= DTW(\varphi, t_j) = \\ DTW(s_i, t_j) &= \|s_i - t_j\| + \min \begin{cases} DTW(s_i, t_{j-1}) \\ DTW(s_{i-1}, t_j) \\ DTW(s_{i-1}, t_{j-1}) \end{cases} \end{aligned} \quad (1)$$

where φ is the empty trajectory. The results of the above equations are imported into a new matrix, called the *similarity matrix*, which displays the relative similarities of all of the trajectories (Figure 2). This matrix is a squared symmetric matrix whose dimension depends on the number of trajectories (n).

$$Similarity = \begin{matrix} & \begin{matrix} R & S & T & \dots \end{matrix} \\ \begin{matrix} R \\ S \\ T \\ \vdots \end{matrix} & \begin{bmatrix} 0 & Sim_{R,S} & Sim_{R,T} & \dots \\ Sim_{R,S} & 0 & Sim_{S,T} & \dots \\ Sim_{R,T} & Sim_{S,T} & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{matrix} \Big]_{n \times n}$$

Fig. 2. Relative similarity measure of the trajectories.

2.3 Pattern discovery based on similarity search

The phrase *pattern discovery* is denoted as “*fitting a model to data, revealing structures, or inferring high-level delineation from the dataset*” (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). This sub-section, therefore, intends to discover the most prominent patterns among the trajectories with respect to the commonalities in their movement parameters and contexts. To this aim, the results of similarity measures are used as the basis for discovering patterns among trajectories. For this, it is required to define a reference range in which the trajectories that fit into it will share identical patterns. Normally, such ranges are distinguished by thresholds (ϵ). We calculate the matching thresholds as one-quarter of the maximum standard deviation of the trajectory similarities in each movement parameter or context, which is also confirmed by Chen and his colleagues (Chen, Özsu, & Oria, 2005). Next, a reference trajectory is selected, and trajectories for which their distances are equal and/or smaller than ϵ fit into the reference range and demonstrate similar patterns.

3. Implementation and Results

After describing the concept of context and providing theories for contextualizing similarity measure and pattern discovery among trajectories, it is time to assess the theories by applying them on a real dataset. We first introduce the case study. Second, we apply the DTW method on the dataset and measure the relative similarities of trajectories. Third, we mine multifarious patterns among trajectories in accordance with their similarity values.

3.1 Data

To evaluate the proposed theory, DTW is applied on 324 real commercial airplane trajectories (Flightaware, 2016) by considering the weather conditions (Aviation Weather Center, 2016) during the flights. Studying airplane movements and measuring their trajectory similarities in different contexts not only enables the extraction of movement-behavior patterns of airplanes but also helps the decision makers in better planning of the flights and predicting their trajectories according to the contexts (Hurter et al., 2014). Commercial airplanes normally fly through predefined routes, called *airways*. An airway connects two locations and is defined with segments within a specific altitude and corridor width and between fixed geographic coordinates, called *waypoints*. In this case study, an airway is chosen from the eastern part of the United States of America, between the Orlando international airport in Orlando, Florida (ICAO code: KMCO), and the John F. Kennedy international airport in New York City, New York (ICAO code: KJFK), as depicted in Figure 3.

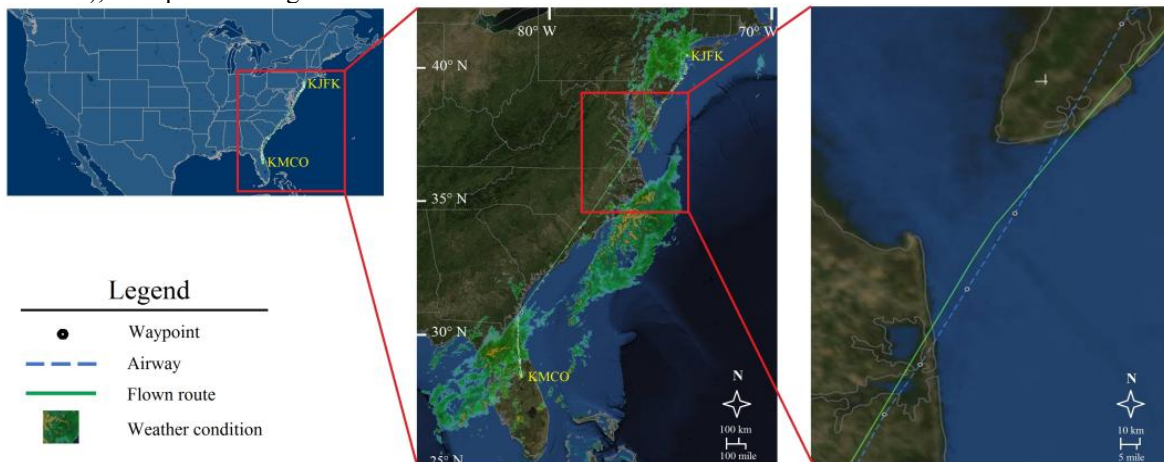


Fig. 3. The study area between KMCO and KJFK airports. (Flightaware, 2016)

3.2 Similarity search based on underlying context

First, the DTW method measures the distance between several trajectories by using spatial data and contextual information in a multi-dimensional space. More specifically, multifarious implementations of (i) positional data (i.e., airplanes' latitude, longitude, and altitude), (ii) internal context information (i.e., airplane's speed and heading), and (iii) external context information (i.e., wind speed and direction) are conducted individually and collaboratively to measure the distance of trajectories in each implementation.

To more clarify how internal and external contexts can affect the similarity measure results, four sample trajectories (Figure 4a and b) along with the weather condition during the flights (Figure 4c) are chosen from the dataset and numerically analyzed. Trajectories K and L are generated via Airbus airplanes, while trajectories M and N are generated by Boeing and McDonnell Douglas airplanes. A 4-by-4 similarity matrix is generated for each examination, where the elements represent the relative similarity values. By applying DTW method, the distance (similarity) of trajectories K and L (Figure 4 red and green lines) at two- and three-spatial dimensions are measured as 7.65 and 2.56, respectively (Figure 5a, 5b). By incorporating airplane's speed and heading parameters, this value decreases to 1.89 (Figure 5c). Given the far similarity distance of these two trajectories in wind speed and direction (42.07) (Figure 5d), the ultimate similarity of trajectories K and L increased to 13.37 (Figure 5e). In contrast, two trajectories M and N (Figure 5 blue and black lines) had high two- and three-spatial similarity values of 45.27 and 15.10, respectively (Figure 5a, 5b). By incorporating airplanes' speed and heading parameters, this value decreased to 9.62 (Figure 5c). Considering the commonalities of these trajectories in wind speed and direction (30.60) (Figure 5d), the final relative similarity of trajectories M and N was increased to 15.61 (Figure 5e). Comparing these two procedures, for trajectories K and L, the similarity value was increased from 7.65 (in 2D) to 13.37 (for all dimensions); while, for trajectories M and N, this value was decreased from 45.27 (in 2D) to 15.61 (for all dimensions). This example reveals that how internal and external contexts can positively/negatively influence the similarity values of trajectories.

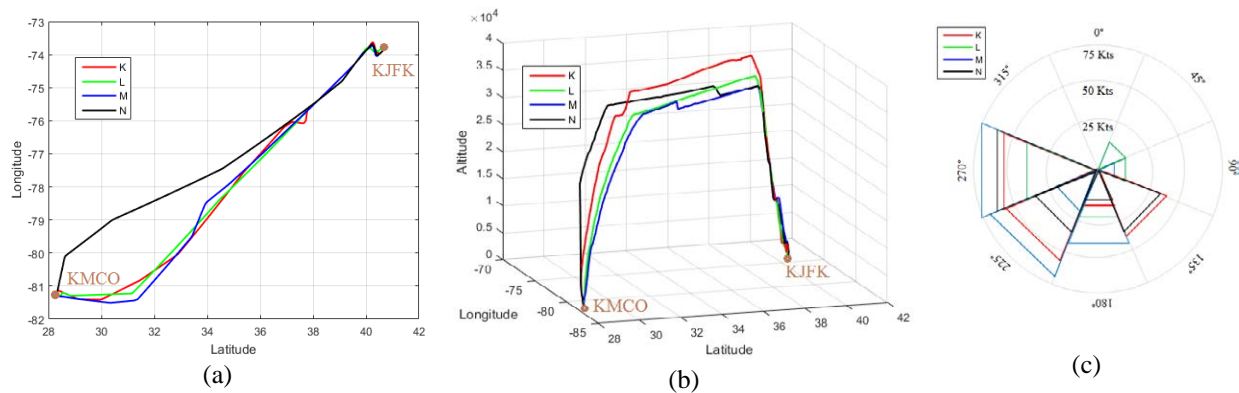


Fig. 4. Representation of four trajectories between KMCO and JFK airports: (a) Airplanes latitude and longitude; (b) Airplanes latitude, longitude, and altitude; (c) Wind (direction and magnitude) rose diagram.

	K	L	M	N		K	L	M	N		K	L	M	N		K	L	M	N
K	0	7.65	13.13	39.36	K	0	2.56	4.38	13.13	K	0	1.89	2.90	8.11	K	0	42.07	28.54	0.11
L	7.65	0	12.21	45.46	L	2.56	0	4.07	15.16	L	1.89	0	2.71	9.52	L	42.07	0	11.74	42.82
M	13.13	12.21	0	45.27	M	4.38	4.07	0	15.10	M	2.90	2.71	0	9.62	M	28.54	11.74	0	30.60
N	39.36	45.46	45.27	0	N	13.13	15.16	15.10	0	N	8.11	9.52	9.62	0	N	0.11	42.82	30.60	0
	(a)					(b)					(c)					(d)			
K	0	13.37	10.22	5.83	K	0	13.37	10.22	5.83	K	0	13.37	10.22	5.83	K	0	13.37	10.22	5.83
L	13.37	0	5.29	19.04	L	13.37	0	5.29	19.04	L	13.37	0	5.29	19.04	L	13.37	0	5.29	19.04
M	10.22	5.29	0	15.61	M	10.22	5.29	0	15.61	M	10.22	5.29	0	15.61	M	10.22	5.29	0	15.61
N	5.83	19.04	15.61	0	N	5.83	19.04	15.61	0	N	5.83	19.04	15.61	0	N	5.83	19.04	15.61	0
	(e)					(e)					(e)					(e)			

Fig. 5. Relative similarity values of four trajectories: (a) Latitude and longitude (2D); (b) Latitude, longitude, and altitude (3D); (c) Latitude, longitude, altitude, airplane speed, and airplane heading; (d) Wind speed and wind direction; (e) All the previous dimensions together.

3.3 Pattern discovery based on similarity search results

It is feasible to investigate the added values and the role of internal and external contextual information in discovering patterns individually and collaboratively. For this purpose, in this sub-section the results of similarity searches are utilized for discovering the multi-dimensional movement patterns of the airplanes' trajectories. To this aim, a reference trajectory is selected; the one that has the least deviation from the airway and the minimum difference between the scheduled and flown flight durations. Then, the implementations are conducted based on this reference trajectory.

The first examination discovers the flights that are close in their duration (i.e., the difference between take-off and landing times). The discovered trajectories are illustrated in Figures 6a (in dark green) in the whole dataset (in light green) based on a reference trajectory (in red). The second examination is dedicated to trajectories that share similar patterns in terms of altitude, time, and airplane's ground speed in Figure 6b. The computed counts of these trajectories are summarized in Table 1.

It is desired to discover patterns of four airplane models in the dataset as the third experiment. To reach this goal, one reference trajectory is selected for each model, and the corresponding spatial patterns of 324 trajectories in three dimensions are extracted. Figures 6c, 6d, 6e, and 6f illustrate these patterns for Boeing, Airbus, McDonnell Douglas, and Embraer, respectively, and the computed counts are summarized in Table 1 (Experiments 3a-3d). The outcome suggests that the Airbus trajectories are the stabilized trajectories in this test because they have the least distances to the reference pattern.

The rest of this sub-section is dedicated to three settings of spatial and contextual information. In this context, the fourth and fifth experiments discover the combined patterns of spatial and internal context (i.e., airplane's speed and heading), and spatial and external context (i.e., wind speed and direction), respectively. The former examination discovers 254 trajectories and the later examination distinguishes 193 trajectories that share similar patterns. Finally, for the combined context assessment, the sixth experiment accounts all the previous variables and discovers 175 multi-dimensional similar patterns among the dataset. The results of the above implementations demonstrate a large variation in the number of discovered trajectories despite the fact that their reference trajectory is the same. The results ascertain the robustness of this technique in quantifying the commonalities of trajectories and discovering movement patterns with overall 80% accuracy.

Table 1. Multi-dimensional pattern discovery of trajectories.

Experiment	Spatial			Temporal	Internal Context		External Context		No. of discovered trajectories
	Latitude	Longitude	Altitude	Flight duration	Airplane's speed	Airplane's heading	Wind speed	Wind direction	
1	✓	✓	-	✓	-	-	-	-	281 out of 324
2	-	-	✓	✓	✓	✓	-	-	267 out of 324
3a	✓	✓	✓	-	-	-	-	-	35 out of 44
3b	✓	✓	✓	-	-	-	-	-	177 out of 218
3c	✓	✓	✓	-	-	-	-	-	37 out of 44
3d	✓	✓	✓	-	-	-	-	-	13 out of 18
4	✓	✓	✓	-	✓	✓	-	-	254 out of 324
5	✓	✓	✓	-	-	-	✓	✓	193 out of 324
6	✓	✓	✓	-	✓	✓	✓	✓	175 out of 324

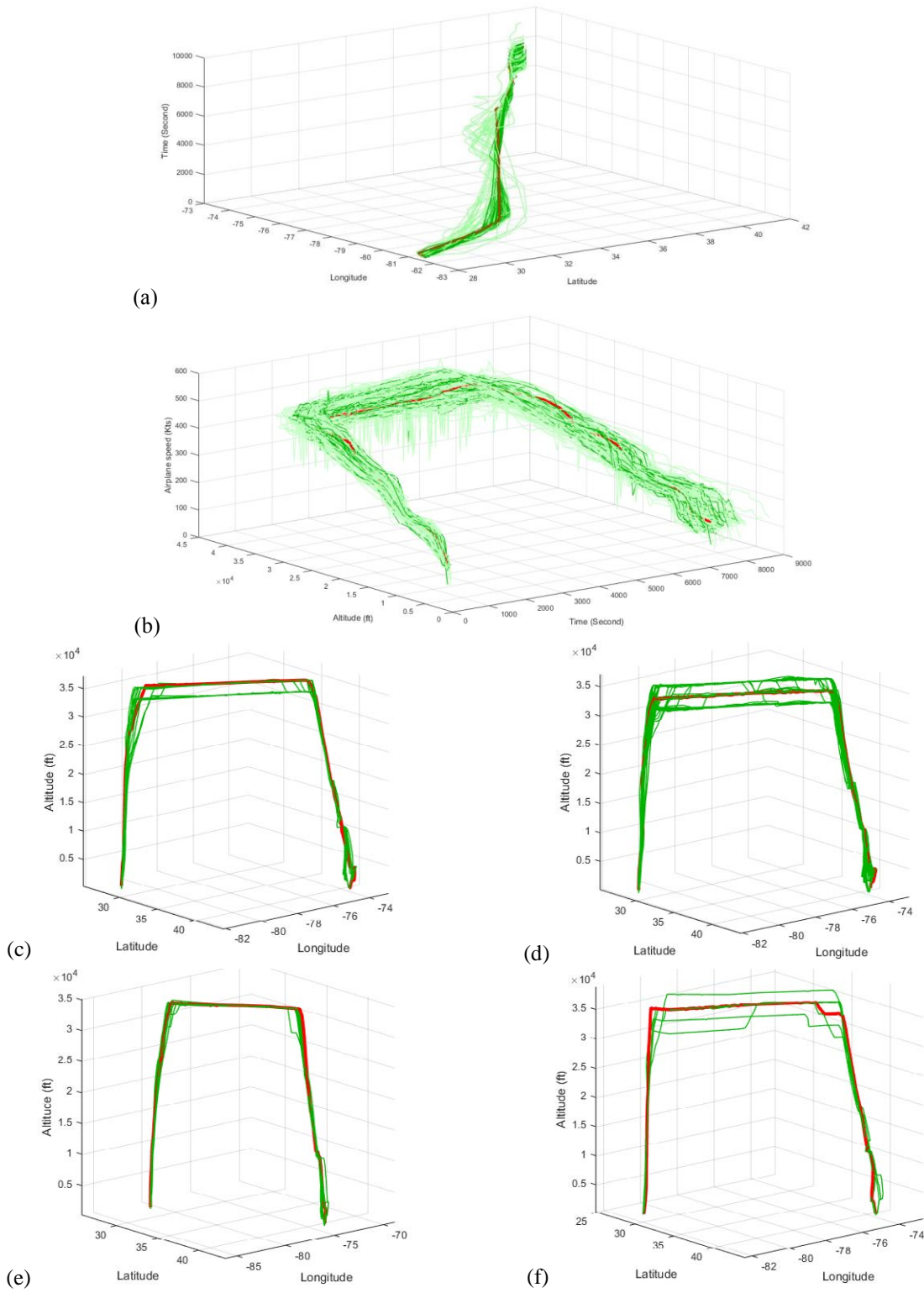


Fig. 6. Discovered patterns (in dark green) from the dataset (in light green) based on the reference trajectory (in red): (a) latitude, longitude, and time; (b) Altitude, airplane's ground speed, and time; (c) Boeing pattern; (d) Airbus pattern; (e) McDonnell Douglas pattern; (f) Embraer pattern.

4. Conclusion

This paper provided new insights into the similarity analysis and pattern discovery of trajectories based on contextual information. To this end, the DTW method was employed for assessing the similarities of trajectories, by including not only their spatial footprints but also a notion of internal and external contexts. The effectiveness of the theory was examined in several experiments on real commercial airplanes dataset. The results demonstrated that the DTW method can be successfully applied in similarity analysis of trajectories and can help to provide pattern discovery in large datasets. Furthermore, the results underpin the fact that the movements of point objects are highly affected by both internal and external contexts in positive and negative manners.

References

- Andrienko, G., Andrienko, N., & Heurich, M. (2011). An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, 25(9), 1347-1370. doi: 10.1080/13658816.2011.556120
- Aviation Weather Center. (2016). www.aviationweather.gov. Accessed 10 December 2016.
- Buchin, M., Dodge, S., & Speckmann, B. (2014). Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science*(9), 101-124.
- Chen, L., Özsu, M. T., & Oria, V. (2005). *Robust and fast similarity search for moving object trajectories*. Paper presented at the Proceedings of the 2005 ACM SIGMOD international conference on Management of data.
- Crowley, J. L., Coutaz, J., Rey, G., & Reignier, P. (2002). Perceptual Components for Context Aware Computing. In G. Borriello & L. E. Holmquist (Eds.), *Proceedings UbiComp 2002: Ubiquitous Computing: 4th International Conference Göteborg, Sweden, September 29 – October 1, 2002* (pp. 117-134). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dey, A. K. (2001). Understanding and Using Context. *Personal Ubiquitous Computing*, 5(1), 4-7. doi: 10.1007/s007790170019
- Dodge, S., Weibel, R., & Laube, P. (2011). Trajectory similarity analysis in movement parameter space. *Plymouth, UK: Proceedings of GISRUK*, 27-29.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Flightaware. (2016). www.flightaware.com. Accessed 10 December 2016.
- Hurter, C., Alligier, R., Gianazza, D., Puechmorel, S., Andrienko, G., & Andrienko, N. (2014). Wind parameters extraction from aircraft trajectories. *Computers, Environment and Urban Systems*, 47, 28-43. doi: <http://dx.doi.org/10.1016/j.compenvurbsys.2014.01.005>
- Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105(49), 19052-19059. doi: 10.1073/pnas.0800375105
- Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43-49. doi: 10.1109/TASSP.1978.1163055
- Spaccapietra, S., Parent, C., & Spinsanti, L. (2013). Trajectories and Their Representations. In C. Renso, S. Spaccapietra & E. Zima'nyi (Eds.), *Mobility data : modeling, management, and understanding* (pp. 394). New York, USA: Cambridge University.